

Problem 73: Collecting voles

(✓)

A group of biologists attempts to estimate the magnitude, N , of an island population of voles (*Microtus agrestis*). Accordingly, the biologists capture a random sample of 200 voles, mark them and release them. A second random sample of 200 voles is then taken of which 11 are found to be marked. Show that the probability, p_N , of this occurrence is given by

$$p_N = k \frac{((N - 200)!)^2}{N!(N - 389)!},$$

where k is independent of N .

The biologists then estimate N by calculating the value of N for which p_N is a maximum. Find this estimate.

All unmarked voles in the second sample are marked and then the entire sample is released. Subsequently a third random sample of 200 voles is taken. Using your estimate for N , write down the probability that this sample contains exactly j marked voles, leaving your answer in terms of binomial coefficients.

Deduce that

$$\sum_{j=0}^{200} \binom{389}{j} \binom{3247}{200-j} = \binom{3636}{200}.$$

2000 Paper II

Comments

This is really just an exercise in combinations. (Recall that a permutation is a reordering of a set of objects, and a combination is a selection of a subset from a set.) You assume that you are equally likely to choose any given subset of the same size, so that the probability of a set of specific composition is the number of ways of choosing a set of that composition divided by the total number of ways of choosing any set of the same size. Of course, you are assuming that the voles are indistinguishable, except for the marks made by the biologists.

Maximising a discrete (not a continuous) function of N came up on one of the previous questions: you have to compare adjacent terms.

The numbers look rather bad, though they turn out OK. My instinct would be to do it algebraically first: replace 200 by a and 11 by b , then substitute back at the end. I am sure it will lead to a better understanding of what is going on.

Solution to problem 73

For the second sample, 200 out of N voles are already marked, so p_N is just the number of ways of choosing 11 from 200 and 189 from $N - 200$ divided by the number of ways of choosing 200 from N :

$$p_N = \frac{\binom{200}{11} \binom{N-200}{189}}{\binom{N}{200}} = \frac{\frac{200!}{11!189!} \frac{(N-200)!}{189!(N-389)!}}{\frac{N!}{200!(N-200)!}},$$

so

$$k = \frac{(200!)^2}{11!(189!)^2}.$$

At the maximum value, $p_N \approx p_{N-1}$, i.e.

$$\frac{(N-200)^2}{N(N-389)} \approx 1,$$

which gives $N \approx 200^2/11 \approx 3636$ (just divide 40000 by 11).

At the third sample, there are 389 marked voles and an estimated $3636 - 389 = 3247$ unmarked voles. Hence

$$P(\text{exactly } j \text{ marked voles}) = \frac{\binom{389}{j} \binom{3247}{200-j}}{\binom{3636}{200}}.$$

The final part follows immediately, using

$$\sum_{j=0}^{200} P(\text{exactly } j \text{ marked voles}) = 1.$$

Post-mortem

Adding the Latin name of the species was a nice touch, I thought (not my idea); it adds an air of verisimilitude to the problem.

I suppose that this might be the basis of a method of estimating the population of voles — rather clever really. I don't know how well it works in practice though. The assumption mentioned earlier, that picking one set of voles of size 200 is just as likely as picking any other set, surely relies on perfect mixing of the marked voles, which would be rather difficult to achieve (especially as female voles can be highly territorial).

You might be asking yourself why it was OK to find the maximum by setting $p_N \approx p_{N-1}$. This is a standard method, but of course it only works if the distribution is one-humped, like a normal distribution. An alternative approach would have been to approximate the distributions using Stirling's approximation, which at its most basic is

$$\ln N! \approx N \ln N - N.$$

This gives exactly the same equation as the $p_N \approx p_{N-1}$ method, and shows that the distribution is indeed one-humped.